
Subject: regression with 95% confidence interval
Posted by [gunvicsin11](#) on Mon, 22 Jul 2013 06:15:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

Respected sir,
Is it possible in idl to use regression and will that procedure provide the 95% confidence interval as output.
Please let me know.
Thanking you in advance
sid

Subject: Re: regression with 95% confidence interval
Posted by [Kenneth Bowman](#) on Mon, 22 Jul 2013 13:35:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 2013-07-22 06:15:37 +0000, sid said:

> Respected sir,
> Is it possible in idl to use regression and will that
> procedure provide the 95% confidence interval as output.
> Please let me know.
> Thanking you in advance
> sid

I e-mailed you a routine that will compute confidence limits on the slope and intercept.

Ken Bowman

Subject: Re: regression with 95% confidence interval
Posted by [David Fanning](#) on Mon, 22 Jul 2013 13:48:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

Kenneth Bowman writes:

> I e-mailed you a routine that will compute confidence limits on the
> slope and intercept.

Could you e-mail it to me, too. :-)

Cheers,

David

--

David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: <http://www.idlcoyote.com/>
Sepore ma de ni thue. ("Perhaps thou speakest truth.")

Subject: Re: regression with 95% confidence interval
Posted by [Fabzi](#) on Mon, 22 Jul 2013 15:11:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 07/22/2013 03:35 PM, Kenneth Bowman wrote:
I e-mailed you a routine that will compute confidence limits on the
> slope and intercept.

I am interested in your method too ;-). Could you mail it to me too?

Until now I did following to test significance (on the F-test):

```
; Prepare data
r = RANDOMNUMBERGENERATOR()
nx = 12
x = FINDGEN(nx)+1

b = 2. ; intercept
sl = 0.5 ; slope
ra = r->GetRandomNumbers(nx) ; some random error
y = sl * x + b + ra

; Now the regression
Np = N_ELEMENTS(y) ; number of points
Nv = 1 ; number of variables (1 for simple reg)
Df = Np - Nv - 1 ; Degrees of freedom

r = regress(x, y, FTEST=ftest)

p_e = 2 * (1 - t_pdf(sqrt(ftest), Df)) ; error probability
if p_e lt 0.01 then Print, 'Significant at the 99% interval'
if p_e lt 0.05 then Print, 'Significant at the 95% interval'
```

I am always wondering what people mean when they write "significant". On
the slope? OR according to the f-Test?

Subject: Re: regression with 95% confidence interval

Posted by [Fabzi](#) on Mon, 22 Jul 2013 15:14:11 GMT

[View Forum Message](#) <> [Reply to Message](#)

On 07/22/2013 03:35 PM, Kenneth Bowman wrote: I e-mailed you a routine that will compute confidence limits on the
> slope and intercept.

I am interested in your method ;-). Could you mail it to me too?

Until now I did following to test significance (on the F-test):

```
; Prepare data
r = RANDOMNUMBERGENERATOR()
nx = 12
x = FINDGEN(nx)+1

b = 2. ; intercept
sl = 0.5 ; slope
ra = r->GetRandomNumbers(nx) ; some random error
y = sl * x + b + ra
```

```
; Now the regression
Np = N_ELEMENTS(y) ; number of points
Nv = 1 ; number of variables (1 for simple reg)
Df = Np - Nv - 1 ; Degrees of freedom
```

```
r = regress(x, y, FTEST=ftest)
```

```
p_e = 2 * (1 - t_pdf(sqrt(ftest), Df)) ; error probability
if p_e lt 0.01 then Print, 'Significant at the 99% interval'
if p_e lt 0.05 then Print, 'Significant at the 95% interval'
```

I am always wondering what people mean when they write "significant". On the slope? OR according to the f-Test?

Subject: Re: regression with 95% confidence interval

Posted by [Kenneth Bowman](#) on Mon, 22 Jul 2013 15:57:47 GMT

[View Forum Message](#) <> [Reply to Message](#)

On 2013-07-22 15:11:51 +0000, Fabien said:

```
> On 07/22/2013 03:35 PM, Kenneth Bowman wrote:
> I e-mailed you a routine that will compute confidence limits on the
>> slope and intercept.
>
> I am interested in your method too ;-). Could you mail it to me too?
```

It is a short routine, so I have pasted it below.

The function computes a linear regression and confidence limits on the slope and intercept following the examples in Tamhane and Dunlop, which is an introductory graduate-level statistics book. As always, the robustness of the confidence limits depends on whether the underlying statistics of the process in question satisfy the assumptions that are made in deriving the confidence limits. In this case the predictor (x) is assumed to be known with no error, and the error in the predictand (y) is assumed normal. It wouldn't hurt to look at Tamhane and Dunlop, for more details.

Cheers, Ken

```
FUNCTION REGRESSION_STATISTICS_KPB, x, y, $
  CI      = ci,      $
  DOUBLE  = double, $
  VERBOSE = verbose

;+
; Name:
;   REGRESSION_STATISTICS_KPB
; Purpose:
;   This program computes the regression coefficients a and b and
their confidence limits
;   for simple bivariate linear regression
;
;    $y = a + b*x + \text{eps}$ 
;
;   The calculations follow the exposition in Chap. 10 of Tamhane and
Dunlop (2000).
; Calling sequence:
;   stats = REGRESSION_STATISTICS_KPB(x, y[, ci])
; Inputs:
;   x      : Values of the independent variables.
;   y      : Values of the independent variables.
; Output:
;   stats  : Data structure containing various statistics and
confidence limits.
; Keywords:
;   CI     : Optional confidence limit (%). If not set, a default
value of 95% is used.
;   VERBOSE : If set, print the statistics and confidence intervals.
; Author and history:
;   Kenneth P. Bowman. 2007-11-01.
;-
```

```

IF (N_PARAMS() NE 2) THEN MESSAGE, 'Incorrect number of arguments.'
;Check number of arguments

IF (N_ELEMENTS(ci) EQ 0) THEN ci = 95.0D0
;Default confidence limit to calculate

n = N_ELEMENTS(x)
;Number of data points
IF (N_ELEMENTS(y) NE n) THEN $
;Check array sizes
MESSAGE, 'The number of x and y values must be equal.'

b = (REGRESS(x, y, CONST = a, YFIT = yfit, FTEST = f_stat, $
;Compute regression and extract value of b
CORRELATION = r, CHISQ = chisq, DOUBLE = double))[0]

s = SQRT(TOTAL((yfit - y)^2)/(n-2))
;Standard deviation of the residuals
x_mean = MEAN(x)
;Mean of the x values
y_mean = MEAN(y)
;Mean of the y values
S_xx = TOTAL((x - x_mean)^2)
;Sum of the squared deviations of x from the mean
S_yy = TOTAL((y - y_mean)^2)
;Sum of the squared deviations of y from the mean
S_xy = TOTAL((x - x_mean)*(y - y_mean))
;Sum of the products of the deviations of x and y from their
means
SE_a = s*SQRT(TOTAL(x^2)/(n*S_xx))
;Standard error of a
SE_b = s/SQRT(S_xx)
;Standard error of b
alpha = (1.0D0 - 0.01D0*ci)/2.0D0
;Level for t-test
t_stat = T_CVF(alpha, n-2)
;Compute t-statistic

IF KEYWORD_SET(verbose) THEN BEGIN
PRINT, 'Intercept (a)      : ', a
PRINT, 'Slope (b)         : ', b
PRINT, 'Correlation coefficient r : ', r
PRINT, 'r^2                : ', r^2
PRINT, 'F-statistic         : ', f_stat
PRINT, 'Chi-square statistic   : ', chisq
PRINT, 'n                    : ', n
PRINT, 'Mean of x              : ', x_mean

```

```

PRINT, 'Mean of y          : ', y_mean
PRINT, 'S.D. of residuals : ', s
PRINT, 'S_xx              : ', S_xx
PRINT, 'S_yy              : ', S_yy
PRINT, 'S_xy              : ', S_xy
PRINT, 'S.E. of a         : ', SE_a
PRINT, 'S.E. of b         : ', SE_b
PRINT, 'Confidence limit   : ', ci, '%'
PRINT, 'Level for t-test   : ', alpha
PRINT, 't-statistic        : ', t_stat
PRINT, 'Confidence interval for a : ', a, '+/-',
STRTRIM(t_stat*SE_a, 2), ' [', a - t_stat*SE_a, ', ', a + t_stat*SE_a,
']'
PRINT, 'Confidence interval for b : ', b, '+/-',
STRTRIM(t_stat*SE_b, 2), ' [', b - t_stat*SE_b, ', ', b + t_stat*SE_b,
']'
ENDIF

RETURN, {a          : a,          $
;Y-intercept
      b          : b,          $
;Slope
      r          : r,          $
;Correlation coefficient r
      yfit       : yfit,       $
;Fitted values at x
      f_stat     : f_stat,     $
;F-statistic
      chisq      : chisq,      $
;Chi-squared statistic
      n          : n,          $
;Number of points
      x_mean     : x_mean,     $
;Mean of the x values
      y_mean     : y_mean,     $
;Mean of the y values
      s          : s,          $
;Standard deviation of the residuals
      S_xx       : S_xx,       $
;Sum of the squared deviations of x from its mean
      S_yy       : S_yy,       $
;Sum of the squared deviations of y from its mean
      S_xy       : S_xy,       $
;Sum of the products of the deviations of x and y from their means
      SE_a       : SE_a,       $
;Standard error of a
      SE_b       : SE_b,       $
;Standard error of b

```

```

        ci      : ci,      $
;Requested confidence interval (%)
        alpha   : alpha,   $
;Level for t-test
        t_stat  : t_stat,  $
;t-statistic value
        ci_a    : t_stat*SE_a, $
;Confidence interval for a is (a +/- ci_a)
        ci_b    : t_stat*SE_b }
;Confidence interval for b is (b +/- ci_b)

```

END

Subject: Re: regression with 95% confidence interval
 Posted by [ftian2012](#) on Mon, 06 Mar 2017 10:04:20 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Monday, July 22, 2013 at 5:57:47 PM UTC+2, KenBowman wrote:

```

> On 2013-07-22 15:11:51 +0000, Fabien said:
>
>> On 07/22/2013 03:35 PM, Kenneth Bowman wrote:
>> I e-mailed you a routine that will compute confidence limits on the
>>> slope and intercept.
>>
>> I am interested in your method too ;-). Could you mail it to me too?
>
> It is a short routine, so I have pasted it below.
>
> The function computes a linear regression and confidence limits on the
> slope and intercept following the examples in Tamhane and Dunlop, which
>
> robustness of the confidence limits depends on whether the underlying
> statistics of the process in question satisfy the assumptions that are
>
> is assumed to be known with no error, and the error in the predictand
>
> for more details.
>
> Cheers, Ken
>
>
> FUNCTION REGRESSION_STATISTICS_KPB, x, y, $
>   CI      = ci,      $
>   DOUBLE  = double, $
>   VERBOSE = verbose
>
>

```

```

> ;+
> ; Name:
> ;   REGRESSION_STATISTICS_KPB
> ; Purpose:
> ;   This program computes the regression coefficients a and b and
> their confidence limits
> ;   for simple bivariate linear regression
> ;
> ;    $y = a + b \cdot x + \text{eps}$ 
> ;
> ;   The calculations follow the exposition in Chap. 10 of Tamhane and
> Dunlop (2000).
> ; Calling sequence:
> ;   stats = REGRESSION_STATISTICS_KPB(x, y[, ci])
> ; Inputs:
> ;   x   : Values of the independent variables.
> ;   y   : Values of the independent variables.
> ; Output:
> ;   stats : Data structure containing various statistics and
> confidence limits.
> ; Keywords:
> ;   CI   : Optional confidence limit (%). If not set, a default
> value of 95% is used.
> ;   VERBOSE : If set, print the statistics and confidence intervals.
> ; Author and history:
> ;   Kenneth P. Bowman. 2007-11-01.
> ;-
>
> IF (N_PARAMS() NE 2) THEN MESSAGE, 'Incorrect number of arguments.'
>   ;Check number of arguments
>
> IF (N_ELEMENTS(ci) EQ 0) THEN ci = 95.0D0
>   ;Default confidence limit to calculate
>
> n = N_ELEMENTS(x)
>   ;Number of data points
> IF (N_ELEMENTS(y) NE n) THEN $
>   ;Check array sizes
>   MESSAGE, 'The number of x and y values must be equal.'
>
> b = (REGRESS(x, y, CONST = a, YFIT = yfit, FTEST = f_stat, $
>   ;Compute regression and extract value of b
>   CORRELATION = r, CHISQ = chisq, DOUBLE = double))[0]
>
> s   = SQRT(TOTAL((yfit - y)^2)/(n-2))
>   ;Standard deviation of the residuals
> x_mean = MEAN(x)
>   ;Mean of the x values

```

```

> y_mean = MEAN(y)
>      ;Mean of the y values
> S_xx = TOTAL((x - x_mean)^2)
>      ;Sum of the squared deviations of x from the mean
> S_yy = TOTAL((y - y_mean)^2)
>      ;Sum of the squared deviations of y from the mean
> S_xy = TOTAL((x - x_mean)*(y - y_mean))
>      ;Sum of the products of the deviations of x and y from their
> means
> SE_a = s*SQRT(TOTAL(x^2)/(n*S_xx))
>      ;Standard error of a
> SE_b = s/SQRT(S_xx)
>      ;Standard error of b
> alpha = (1.0D0 - 0.01D0*ci)/2.0D0
>      ;Level for t-test
> t_stat = T_CVF(alpha, n-2)
>      ;Compute t-statistic
>
> IF KEYWORD_SET(verbose) THEN BEGIN
>   PRINT, 'Intercept (a)      : ', a
>   PRINT, 'Slope (b)         : ', b
>   PRINT, 'Correlation coefficient r : ', r
>   PRINT, 'r^2               : ', r^2
>   PRINT, 'F-statistic       : ', f_stat
>   PRINT, 'Chi-square statistic : ', chisq
>   PRINT, 'n                 : ', n
>   PRINT, 'Mean of x         : ', x_mean
>   PRINT, 'Mean of y         : ', y_mean
>   PRINT, 'S.D. of residuals : ', s
>   PRINT, 'S_xx              : ', S_xx
>   PRINT, 'S_yy              : ', S_yy
>   PRINT, 'S_xy              : ', S_xy
>   PRINT, 'S.E. of a         : ', SE_a
>   PRINT, 'S.E. of b         : ', SE_b
>   PRINT, 'Confidence limit   : ', ci, '%'
>   PRINT, 'Level for t-test   : ', alpha
>   PRINT, 't-statistic       : ', t_stat
>   PRINT, 'Confidence interval for a : ', a, '+/-',
>   STRTRIM(t_stat*SE_a, 2), ' [', a - t_stat*SE_a, ', ', a + t_stat*SE_a,
>   ']'
>   PRINT, 'Confidence interval for b : ', b, '+/-',
>   STRTRIM(t_stat*SE_b, 2), ' [', b - t_stat*SE_b, ', ', b + t_stat*SE_b,
>   ']'
> ENDIF
>
> RETURN, {a      : a,      $
>      ;Y-intercept
>      b      : b,      $

```

```

> ;Slope
>   r      : r,      $
> ;Correlation coefficient r
>   yfit    : yfit,    $
> ;Fitted values at x
>   f_stat   : f_stat,   $
> ;F-statistic
>   chisq    : chisq,    $
> ;Chi-squared statistic
>   n        : n,        $
> ;Number of points
>   x_mean   : x_mean,   $
> ;Mean of the x values
>   y_mean   : y_mean,   $
> ;Mean of the y values
>   s        : s,        $
> ;Standard deviation of the residuals
>   S_xx     : S_xx,     $
> ;Sum of the squared deviations of x from its mean
>   S_yy     : S_yy,     $
> ;Sum of the squared deviations of y from its mean
>   S_xy     : S_xy,     $
> ;Sum of the products of the deviations of x and y from their means
>   SE_a     : SE_a,     $
> ;Standard error of a
>   SE_b     : SE_b,     $
> ;Standard error of b
>   ci       : ci,       $
> ;Requested confidence interval (%)
>   alpha    : alpha,    $
> ;Level for t-test
>   t_stat   : t_stat,   $
> ;t-statistic value
>   ci_a     : t_stat*SE_a, $
> ;Confidence interval for a is (a +/- ci_a)
>   ci_b     : t_stat*SE_b }
> ;Confidence interval for b is (b +/- ci_b)
>
> END

```

Thank you very much KenBowman for sharing your excellent code!
